



ANALYSIS OF LOW-INTENSITY SCINTILLATION SPECTRA

V. Muravsky, S.A. Tolstov

International Sakharov Institute of Radioecology,
Minsk, Belarus

Abstract. The maximum likelihood algorithms for nuclides activities estimation from low intensity scintillation γ -ray spectra have been created. The algorithms treat full energy peaks and Compton parts of spectra, and they are more effective than least squares estimators. The factors that could lead to the bias of activity estimates are taken into account. Theoretical analysis of the problem of choosing the optimal set of initial spectra for the spectrum model to minimize errors of the activities estimation has been carried out for the general case of the N -components with Gaussian or Poisson statistics. The obtained criterion allows to exclude superfluous initial spectra of nuclides from the model. A special calibration procedure for scintillation γ -spectrometers has been developed. This procedure is required for application of the maximum likelihood activity estimators processing all the channels of the scintillation γ -spectrum, including the Compton part. It allows one to take into account the influence of the sample mass density variation. The algorithm for testing the spectrum model adequacy to the processed scintillation spectrum has been developed. The algorithms are realized in Borland Pascal 7 as a library of procedures and functions. The developed library is compatible with Delphi 1.0 and higher versions. It can be used as the algorithmic basis for analysis of highly sensitive scintillation γ - and β -spectrometric devices.

Estimators for nuclide activities in low-intensity scintillation spectrometry*Calculation of the nuclides activities estimates by the maximum likelihood method*

Ionizing radiation has a random character, therefore the number of counts $S(i)$ in the i -th channel of the spectrum S with a finite accumulation time T is a random variable with the mean value $\nu(i)$. A distribution of counts in each of the spectrum channels is a Poisson or polynomial distribution. A polynomial distribution takes place if spectrum is accumulated so long as the spectrum integral less than a definite value.

In a case of the low-intensity radiation one may write the mean value $\nu(i)$ of the number of counts in the i -th spectrum channel as

$$\nu(i) = T \sum_{k=1}^{M+1} A_k f_k(i), \quad (1)$$

where T is the number of nuclides; M is the number of nuclides; $f_k(i)$ ($k \leq M$) is the spectrum of the k -th nuclide, normalized to unit activity and unit time (the "initial spectrum" of the k -th nuclide); A_k ($k \leq M$) is the activity of the k -th nuclide; $f_{M+1}(i)$ is the "initial spectrum" of a background; A_{M+1} is the relative intensity of a background.

The problem of nuclide activity estimation should be solved by the statistical methods. A general approach to the problem of the correct estimation of random distribution parameters was investigated in a number of classical works. It was shown in ref. [1,2] that the best parameter estimates are obtained by the maximum likelihood (ML) method. In case of a Gaussian distribution the ML method is reduced to the well developed least squares method. However, for a Poisson distribution (or other distributions) the maximum likelihood method is optimal [1,2].

Let us suppose that x is a discrete random value (one-dimensional or multidimensional) with a probability function $P(x, Q_1, \dots, Q_M)$, where Q_k ($k = 1, 2, \dots, M$) are the unknown parameters of the distribution. The ML estimates of the parameters Q_k for observed N values $\{x_1, \dots, x_N\}$ are obtained by maximizing the likelihood function L

$$L(x_1, \dots, x_N, Q_1, \dots, Q_M) = \prod_{i=1}^N P(x_i, Q_1, \dots, Q_M) \quad (2)$$

or its logarithm.

Usually, one uses the inverse matrix to the Fisher's information matrix [1,2] as an estimate of the covariance matrix of ML estimates. Elements I_{kl} of the Fisher's information matrix are equal to

$$I_{kl} = -E \left[\frac{\partial^2 \ln L}{\partial Q_k \partial Q_l} \right], \quad (3)$$

where E means averaging.

The logarithm of the likelihood function for a Poisson distribution in spectrum channels is equal to

$$\ln L = \sum_{i=1}^N \left(-T \sum_{k=1}^{M+1} A_k f_k(i) + S(i) \ln \left(T \sum_{k=1}^{M+1} A_k f_k(i) \right) - \ln(S(i)!) \right). \quad (4)$$

The logarithm of the likelihood function for a polynomial distribution in spectrum channels is equal to

$$\ln L = \ln(I!) + \sum_{i=1}^N \left(S(i) \ln \left(T \sum_{k=1}^{M+1} A_k f_k(i) \right) - S(i) \ln I - \ln(S(i)!) \right), \quad (5)$$

$$\text{where } I = \sum_{i=1}^N S(i) = T \sum_{i=1}^N \sum_{k=1}^{M+1} A_k f_k(i). \quad (6)$$

It could be shown, that the activity estimates for both Poisson and polynomial statistics can be found from the set of the $M+1$ equations

$$\sum_{i=1}^N \left(\frac{S(i) f_j(i)}{\sum_{k=1}^{M+1} A_k f_k(i)} - T f_j(i) \right) = 0 \quad (7)$$

where $j = 1, 2, \dots, M+1$.

In general, the set of equations (7) is non-linear, so it should be solved by numerical methods. We proposed in [3] an effective iteration scheme for solution of (7). In every q -th iteration ($q = 1, 2, \dots$), one solves the set of linear equations:

$$\sum_{k=1}^{M+1} A_k T \sum_{i=1}^N \frac{f_k(i) f_j(i)}{D_q(i)} = \sum_{i=1}^N \frac{S(i) f_j(i)}{D_q(i)}, \quad (8)$$

where

$$D_q(i) = \begin{cases} \sum_{k=1}^{M+1} A_k^{(q-1)} f_k(i), & \text{if } q > 1, \\ [S(i) + 1]/T, & \text{if } q = 1 \end{cases}$$

$A_k^{(q-1)}$ are the estimates of the activities, obtained in $q-1$ iteration.

The elements of the Fisher's information matrix are equal to

$$I = T \sum_{i=1}^N \frac{f_k(i) f_j(i)}{D_q(i)}. \quad (9)$$

So, the covariance matrix of ML estimates of the activities can be found by inversion of the matrix (9).

As a rule, in a case of the low-intensity scintillation γ -spectrometry the initial spectra $f_k(i)$ are obtained experimentally during spectrometer calibration with a set of standard samples (sources) of nuclides activities.

However, the initial spectra, obtained in such way, inescapably contain the statistical noise. The presence of the noise leads to a bias of the estimates, obtained from the iteration scheme (8), because the diagonal elements of the matrix of the set of linear equations, solved in each iteration, are biased. To compensate that bias it is necessary to modify the iteration scheme (8) in the following way [4]:

$$\sum_{k=1}^{M+1} A_k T \sum_{i=1}^N \left[\frac{f_k(i, \rho) f_j(i, \rho) - \delta_{kj} d_j(i, \rho)}{D_q(i)} \right] = \sum_{i=1}^N \frac{S(i, \rho) f_j(i, \rho)}{D_q(i)}, \quad (10)$$

where ρ is a mass density of an investigated sample; $\delta_{kj} = \begin{cases} 1, & \text{if } k = j \\ 0, & \text{if } k \neq j \end{cases}$,

$d_j(i)$ is the statistical noise variation in j -th initial spectrum ("variation" of the j -th initial spectrum),

$$D_q(i) = \begin{cases} 0.5 \cdot [P_q(i-1) + P_q(i+1)], & \text{if } 1 < i < N, \\ P_q(i), & \text{if } i = 1 \text{ or } i = N \end{cases}$$

$$P_q(i) = \begin{cases} \sum_{k=1}^{M+1} A_k^{(q-1)} [f_k(i) + T \cdot A_k^{(q-1)} d_k(i)], & \text{if } q > 1, \\ [S(i) + 1]/T, & \text{if } q = 1 \end{cases}$$

The way of calculation of $D_q(i)$ allows to eliminate the correlation between the numerator and the denominator in the terms

$$\frac{f_k(i, \rho) f_j(i, \rho) - \delta_{kj} d_j(i, \rho)}{D_q(i)} \text{ and } \frac{S(i, \rho) f_j(i, \rho)}{D_q(i)}.$$

That correlation being not eliminated gives additional bias of the activities estimates.

The iteration process (10) is exited, if the following condition is true:

$$\left| A_k^{(q)} - A_k^{(q-1)} \right| < \varepsilon \left| A_k^{(q)} \right| + \tau, \text{ for } k = 1, 2, \dots, M+2, \quad (11)$$

where ε is desirable accuracy of the approximation, τ is the small constant added for the compensation of the influence of the floating-point operations precision on the exit from the iteration scheme. Our experience shows that the recommended values of these constants are $\varepsilon = 10^{-6 \dots 8}$, $\tau = 10^{-8 \dots 10}$ (when using the Borland Pascal real types single (4 bytes) or double (8 bytes)).

The covariance matrix of the activities estimates is calculated by the inversion of the Fisher's information matrix I :

$$I_{vj} = T \sum_{i=1}^N \frac{f_v(i, \rho) f_j(i, \rho) - \delta_{vj} d_j(i, \rho)}{D_q(i)}. \quad (12)$$

As far as the information matrix (12) is the matrix of the set of linear equations of the iteration scheme (10), it is convenient to use in (10) the Gauss-Jordan algorithm [6] for solution of the set of linear equations. The Gauss-Jordan algorithm produces a vector of the solutions of a set of linear equations and the inverse matrix to the matrix of a set of equations. Therefore, the estimates of activities and the covariance matrix of these estimates will be calculated simultaneously.

It is necessary to emphasize that the absence of the correlation between neighboring channels of initial spectra is the necessary condition for application of the iteration scheme (10).

Calibration of scintillation spectrometer

The necessary for activity estimation set of the initial spectra $f_k(i, \rho)$ and their variations $d_k(i, \rho)$ are created from the library of the experimentally obtained during spectrometer calibration initial spectra $f_k(i, \rho_m)$ and their variations $d_k(i, \rho_m)$, here ρ is the density of the investigated sample. The $f_k(i, \rho)$ are calculated as:

$$f_k(i, \rho) = \frac{\rho_{m+1} - \rho}{\rho_{m+1} - \rho_m} f_k(i, \rho_m) + \frac{\rho - \rho_m}{\rho_{m+1} - \rho_m} f_k(i, \rho_{m+1}), \quad (13)$$

and accordingly,

$$d_k(i, \rho) = \left(\frac{\rho_{m+1} - \rho}{\rho_{m+1} - \rho_m} \right)^2 d_k(i, \rho_m) + \left(\frac{\rho - \rho_m}{\rho_{m+1} - \rho_m} \right)^2 d_k(i, \rho_{m+1}), \quad (14)$$

where m has to satisfy to the condition $\rho_m \leq \rho < \rho_{m+1}$.

Initial spectra $f_k(i, \rho_m)$ and their variations $d_k(i, \rho_m)$ are calculated during calibration of scintillation spectrometer in the next way:

$$f_k(i, \rho_m) = \frac{1}{A_{km}} \left[\frac{S_k(i, \rho_m)}{T_k} - \frac{S_{Ph}(i, \rho_m)}{T_{Ph}} \right], \quad (15)$$

$$d_k(i, \rho_m) = \frac{1}{A_{km}^2} \left[\frac{S_k(i, \rho_m)}{T_k^2} + \frac{S_{Ph}(i, \rho_m)}{T_{Ph}^2} \right], \quad (16)$$

where $S_k(i, \rho_m)$ is the spectrum of a sample with the mass density ρ_m containing the k -th nuclide; T_k is the accumulation time of $S_k(i, \rho_m)$; A_{km} is known activity of the k -th nuclide in a sample with the mass density ρ_m ; $S_{Ph}(i, \rho_m)$ is the spectrum of a background measured from a sample without nuclides and with the density ρ_m ; T_{Ph} is the accumulation time of $S_{Ph}(i, \rho_m)$.

Spectra $S_k(i, \rho_m)$ and $S_{Ph}(i, \rho_m)$ are accumulated from standard samples during calibration of scintillation spectrometer.

Those standard samples have to answer the following requirements:

- 1) The set of the standard samples should have such values of the density to represent all operating range of the mass density of the investigated samples.
- 2) The standard samples have to be measured in fixed geometry. The geometry is to be identical to the geometry of measurements of the investigated samples. That is, the same beakers have to be used to fill samples in, samples should fill the beakers at the same level, the samples should be homogeneous in volume and they should have homogeneous distribution of nuclides in volume, etc.
- 3) Deviations of the spectrometric scale should be compensated in each spectrum of the standard samples. Our Software includes procedures, which perform compensation of the deviations of spectrometric scale by means of stretching or compression of the processed spectrum. Such stretching or compression of the processed spectrum is done according to position of a well-resolved full-energy peak in a preliminarily accumulated special reference source (for example ^{137}Cs) of γ -radiation.

Choosing of the spectrum model

Under processing the multi-component spectra it is important to choose an adequate spectrum model. If the number of initial spectra in the chosen spectrum model is less than in reality, then the nuclides activity estimates obtained will be biased. Hence, at the first sight, it seems that the chosen model should include as many initial spectra of the nuclides as it is possible. However, the nuclide activity estimation accuracy depends on the number of initial spectra of the nuclides in the spectrum model. And what is more, the accuracy decreases with increasing of the nuclide number included in a model (till the model is true). Thus, one need a criterion for choosing such a model that allows to minimize the errors of the nuclides activities estimates.

If we know *apriory*, that activity of some nuclide does not exceed A_{max} in the investigated samples with given confidence probability P , then it could be shown theoretically, that the criterion of the exclusion of such nuclide (we shall call it "e-nuclide") from the spectrum model looks like:

$$A_{max} < t_p \sqrt{D(\hat{A})}. \quad (17)$$

Here t_p is the quantile of the normal distribution, corresponding to the given confidence probability P ; $D(\hat{A})$ is variance of the maximum likelihood estimate \hat{A} of the activity of the "e-nuclide", obtained from the spectrum model including that "e-nuclide".

The criterion (17) has been obtained from comparison for any nuclide, which is not "e-nuclide", of full errors of maximum likelihood nuclide activity estimate in the spectrum model including "e-nuclide" and in the spectrum model not including "e-nuclide".

This criterion says, that if we know, that activity of some nuclide does not exceed some know threshold, then that nuclide should be included in the spectrum model only if this threshold does not satisfy condition (17). Thus minimal error of activities estimates of other nuclides is provided.

As far as one does not know A_{max} in practice, it should be reasonable to use estimate of activity of corresponding nuclide instead. Corresponding criterion looks like [5]:

$$\hat{A} < t_p \sqrt{D(\hat{A})}. \quad (18)$$

It is could be seen, that obtained criterion very similar to test of statistical hypothesis of equality to zero of nuclide activity.

Test of goodness of fit

After calculating of nuclide activities it should be reasonable to test the goodness of fit of the experimental spectrum with the spectrum model (1). This test could be done by means of χ^2 test with modified χ^2 functional. That modified functional looks like:

$$\chi^2 = \sum_{i=1}^N \frac{\left[S(i) - T \sum_{k=1}^{M+1} A_k f_k(i) \right]^2}{T \sum_{k=1}^{M+1} A_k \left[f_k(i) + A_k T d_k(i) + A_k T g_k^2(i) D \right]}, \quad (19)$$

where $g_k(i)$ is a function, by means of which we may describe approximately the modification of the shape of k -th initial spectrum influenced by the variation of the spectrometric scale; D is a variance of the scale variation factor δ , which is equal to the relative deviation of the width of the channels of the spectrometer from a give value. The scale variation factor δ is proportional to the gain of the photomultiplier tube (PMT) and the spectrometric amplifier in the scintillation spectrometry.

The meaning of δ and $g_k(i)$ could be seen from the next Taylor's series expansion of the initial spectrum of k -th nuclide:

$$f_k(i) = f_k^{(0)}(i \cdot [1 + \delta_k]) = f_k^{(0)}(i) + \frac{df_k^{(0)}(i)}{di} \cdot i \cdot \delta_k + \dots \approx f_k^{(0)}(i) + \delta_k \cdot g_k(i), \quad (20)$$

Here $f_k^{(0)}(i)$ is initial spectra of k -th nuclide with $\delta_k = 0$; δ_k is the scale variation factor in that initial spectrum.

The scale variation factor δ is variate, which depends on the external factors, which change the gain of the PMT and of the spectrometric amplifier. The mean value of δ is equal to zero, and its variance D could be defined from the technical characteristics of the concrete scintillation spectrometer.

Let us give an example of the calculation of D . Suppose it is indicated in the technical documentation of the scintillation spectrometer, that the maximum of the full energy peak of ^{137}Cs is located in the 220-th channel, and the absolute error of this location is 2 channels for the 95% of the confidence probability. Therefore, the absolute error of the δ for the given confidence probability is equal to $2/220 \approx 0,009$. Let's divide this number by 1,96 (the quantile of the normal distribution, corresponding to the confidence probability being equal to 0,95). As a result, we get the value of the standard deviation of the δ , what is equal to 0,0046. After that, we square this value and get D , which is equal to the $2,1 \cdot 10^{-5}$ in the considered case.

As far as A_k in (19) are the nuclides activities estimates, calculated from $S(i)$ by the maximum likelihood method, we can suppose, that the distribution of functional (19) is tended to the χ^2 -distribution with $N-M-1$ degrees of freedom, if the theoretical model (1) of spectrum is adequate to the processed experimental scintillation spectrum.

To provide closeness of the distribution of functional (19) to the χ^2 -distribution it is necessary to join some channels of the processed spectrum. In practice, it is sufficient to make

$$T \sum_{k=1}^{M+1} A_k f_k(i) > 30 \quad (21)$$

for that [1].

Thus, the testing of the goodness of fit of the processed experimental scintillation spectrum with the spectrum model (1) should be executed after nuclides activities estimation by the maximum likelihood method in the next way.

The value of χ^2 is calculated according to (19). The value $\chi^2_{P(N-M-1)}$ is determined from the tables of χ^2 -distribution or is calculated in some way, here P is a given confidence probability and $N-M-1$ is a number of the degrees of freedom.

And if the inequality

$$T < T_P(N-M-1) \quad (22)$$

is satisfied, then the hypothesis about adequacy of the spectrum model to the experimental spectrum is adopted. Otherwise, this hypothesis is turned down and obtained estimates of the nuclides activities are believed wrong. It is necessary to eliminate the reason of the inadequacy of the theoretical model and to repeat accumulation of the spectrum with consequent estimation of the nuclides activities and testing of the adequacy of the spectrum model. The reason of the inadequacy could be, for example, radioactive contamination of the spectrometer's components, or spectrometer's malfunction, or misfit of the shape of the background spectrum, used in theoretical model, or something else.

Function $g_k(i)$ in (19) may be obtained from $f_k(i)$. Indeed, it follows from (20), that

$$g_k(i) = \frac{df_k^{(0)}(i)}{di} \cdot i \approx i [f_k(i+1) - f_k(i)]. \quad (23)$$

How to use developed software for maximum likelihood estimation of nuclides activities

Library of the algorithms

Developed algorithms for maximum likelihood estimation of nuclides activities are realized in Borland Pascal 7.0 computing language. The library of the procedures and functions, realizing developed algorithms, is organized as the set of the following program modules:

- GLOBUNT.PAS is the library of the auxiliary procedures and functions.
- MATHUNIT.PAS is the library of the auxiliary mathematical function and procedures.
- POLINUNIT.PAS is the library of the functions for processing of polynomials. Where are functions for polynomial spectrum approximation by the least squares method here.
- SPECUNIT.PAS is the library of the functions and procedures for working with the spectra. The structure *TSpectrum* with spectrum data, functions of spectrum saving to disk and spectrum loading from disk are defined here.
- FORMSUNT.PAS is the library of the functions and procedures for working with initial spectra of the nuclides and variations of the initial spectra. There are the function of the spectrometer calibration *MakeNuclideBase* and the function of the spectrometric data preparation *PrepareSpectrometricData* in this module.
- MLMUNIT.PAS is the library of the functions of the nuclides activities estimation by the maximum likelihood method. The functions *MAXLIKMETHOD_WITH_SELECTION*, *MAXLIKMETHOD* and *ChiSqr* are here.

Developed library has been tested for compatibility with Delphi 3.0 and Borland Pascal 7.0. To use developed library in Borland Pascal 7.0 programs you must set DOS in conditional defines.

Testing of the developed program modules

We have developed the programs MLMTEST.EXE, MKSPEC.EXE and VIEWSPEC.EXE for testing of the developed program modules. It is also an example of the usage of the developed program modules. Executable file of the programs and source-code are in the main directory of the Software.

The program MLMTEST.EXE has been optimized for using in the Software of the scintillation γ -spectrometers as a module for the nuclides activities estimation by the ML method. The program is used for the calculation of the spectrometric scale parameters, for the spectrometer calibration and for the estimation of the nuclides activities from the spectra of investigated samples by the ML method.

The program MKSPEC.EXE has been developed for simulation of the γ -spectra of the mixture of nuclides ^{137}Cs , ^{40}K , ^{228}Ra and ^{226}Th . The input data of the program are the spectrum accumulation time and the activities of the nuclides ^{137}Cs , ^{40}K , ^{228}Ra and ^{226}Th . The initial spectra of the nuclides ^{137}Cs , ^{40}K , ^{228}Ra , ^{226}Th and background, used for the spectrum simulation, was obtained from the samples with the density 1.0 g/cm^3 in Marinelli's beaker of 0.5 liter on the γ -spectrometer with NaI(Tl) scintillator (diameter 40 mm and height 40 mm) and 1024-channel analyzer. The files with the initial spectra, used for the spectrum simulation, are found in the directory MKSPEC in files CS137, K40, RA228, TH226 and PHON. The size of the each file is 1024×4 bytes (1024 channels and four bytes (single) for each channel).

The program VIEWSPEC.EXE has been developed to view the spectra, stored in files in the format used by the developed programs and program modules.

Current application of the developed software

Developed program modules are used in production-run equipment: scintillation β - γ -spectrometer EL 1315 with CsI(Na) γ -scintillator (diameter 63 mm and height 63 mm) and β - γ -

radiometer EL 1311 with CsI(Na) γ -scintillator (diameter 150 mm and height 40 mm) and 1024-channel analyzers and radiometer RUG91M with NaI(Tl) scintillator (diameter 40 mm and height 40 mm) and 256-channel analyzer.

Some remarks on application of the developed algorithms

Using the developed algorithms one should take into account the following:

1. The algorithm of the activities estimation were developed on the assumption of the Poisson statistics of counts in spectrum channels. Application of spectra with other statistics, for example, the smoothed spectra, will always lead to the activities estimates with lower accuracy.
2. Another assumption is an absence of the correlation between the statistical noise components of the initial spectra of the nuclides, a background and investigated sample spectrum, used in activity estimation. Therefore, activity estimates, obtained from the spectrum used in spectrometer calibration or from the spectrum being equal to the background spectrum used in the activity estimation, are not correct.

Therefore, while checking the work of the algorithm for activity estimation from a background spectrum, one should have two spectra of a background independently accumulated. One spectrum should be used as the spectrum of a background, the other spectrum should be used as a spectrum of an investigated sample with unknown nuclides activities.

If it is necessary to check the work of the algorithm in activity estimation from the standard sample, used for the spectrometer calibration, one should use separately accumulated spectrum of that standard sample and separately accumulated spectrum of a background.

3. The developed algorithm of the spectrum model choosing based on the elaborated statistical criteria could be applied in expert systems for nuclides identification. This is because the result of this algorithm execution is a list of the identified nuclides with estimated activities and their errors and the result of the test of goodness of fit.

REFERENCES

- [1] JANOSSY, L., Theory and practice of the evaluation of measurements (Oxford University Press, 1965).
- [2] EADIE, W.T. , DRYARD, D., JAMES, F.E., ROOS, M., SADOULET, B., Statistical methods in experimental physics (North-Holland Publishing Company Amsterdam London, 1971).
- [3] MURAVSKY, V.A., TOLSTOV, S.A., KHOLMETSII, A.L., Comparison of the least squares and the maximum likelihood estimators for gamma-spectrometry, Nucl. Instr. and Meth. in Phys. Res., **B145** (1998) 573-577.
- [4] MURAVSKY, V.A., TOLSTOV, S.A., Algorithms of nuclides activities estimation by the maximum likelihood method, Application of semiconductor detectors in nuclear physics: Abstracts of 5th International Conference (Riga, Latvia, May 18-22) 1998, p. 102.
- [5] MURAVSKY, V.A., TOLSTOV, S.A., Statistical criterion for choosing of optimal model of ionizing radiation spectra, Application of semiconductor detectors in nuclear physics: Abstracts of 5th International Conference (Riga, Latvia, May 18-22) 1998, p. 103.
- [6] PRESS, W.H., FLANNERY, B.P., VETTERLING, W.T., Numerical recipes: the art of scientific computing (Cambridge University Press, 1986), 206-208.